Uncertainty Associated with Microbiological Analysis

1. Introduction

1.1. There are only two absolute certainties in life: death and taxes! Whatever task we undertake, no matter how menial or how sophisticated, we are faced with a lack of certainty in the outcome! It is therefore essential to have a common understanding of what is meant by uncertainty in relation to our specific tasks in defining BPMM.

1.2. In microbiological laboratory practice, we can identify many causes of variability, for instance:

1.2.1. The ability of an isolate to give typical reactions on a diagnostic medium;

1.2.2. The use of the incorrect ingredients in a culture medium;

1.2.3. The consequence of changing brands of commercial media;

1.2.4. Use of non-standard conditions in the preparation, sterilisation and use of a culture medium;

1.2.5. Equipment and human errors in weighing, dispensing, pipetting and other laboratory activities;

1.2.6. The tolerance applied to the shelf life of test reagents;

1.2.7. The relative skill levels of different technicians;

1.2.8. The relative well-being of any technician who is undertaking analyses;

1.2.9. and so on, and so on …. *ad infinitum!*

1.3. These are but a few trite examples of biological, instrumental and personal bias that affect the accuracy, precision and hence the uncertainty of microbiological tests; a situation that constantly faces scientists involved in laboratory management.

1.4. To interpret properly the results obtained using any analytical procedure, whether physical, chemical or biological, requires careful consideration of the diverse sources of actual or potential error associated with the results obtained. Any analytical result is influenced by a complex of three major error groups:

1.4.1. *Random errors*, associated with the original sample matrix, the analytical (test) sample, the culture media, etc;

1.4.2. Inherent *systematic errors* associated with the analytical procedure; and

1.4.3. Modification of the systematic errors due to a particular laboratory's environment and equipment together with individual analysts' personal traits in carrying out the test procedure.

1.5. **Accuracy and Precision**

1.5.1. Accuracy is a qualitative concept (VIM, 1993). In simple terms, accuracy can be defined as the correctness of a result, relative to an expected outcome; whilst precision is a measure of the variability of test results.

1.5.2. Accuracy is defined (ISO3534-2:2003) as "*the closeness of agreement between a test result or a measurement result and the true value*." Accuracy is a combination of trueness and precision (a combination of random components and systematic error or bias components). This differs from the definition given by VIM (1993): "*the closeness of agreement between the result of a measurement and a true value of a measurand*".

1.5.3. "***Accuracy***" is essentially "absence of error"; the more accurate a result the lower the associated error of the test. It is important to note that the term "accuracy " applies only to results and can not be applied to methods, equipment, laboratories or other general matters.

1.5.4. "***Trueness***" is defined (ISO, 2003) as, "the closeness of agreement between the average value obtained from a large series of test results and an accepted reference value".

1.5.5. *Trueness* is equivalent to an absence of "***bias***", which is the difference between the expectation of the test results and an accepted reference value and is a measure of total systematic, but not random, error.

1.5.6. *Trueness*, unlike *accuracy*, may correctly be contrasted with *precision.*

1.6. "***Precision"*** is defined as the closeness of agreement between independent test results obtained under stipulated conditions.

1.6.1. *Precision* depends only on the distribution of random errors and does not relate to a true value or a specified value.

1.6.2. The *measure of precision* is expressed usually in terms of imprecision and computed as a standard deviation of the test results.

1.6.3. Lower precision is reflected by a larger standard deviation.

1.6.4. *Independent test results* means results obtained in a manner not influenced by any previous results on the same or similar test object.

1.6.5. *Quantitative measures of precision* depend critically on the stipulated conditions. Repeatability and reproducibility conditions are particular sets of extreme stipulated conditions (ISO 3534: 3.14).

1.7. Fig 1 illustrates schematically the relationships between trueness, accuracy, precision and uncertainty (AMC, 2003).

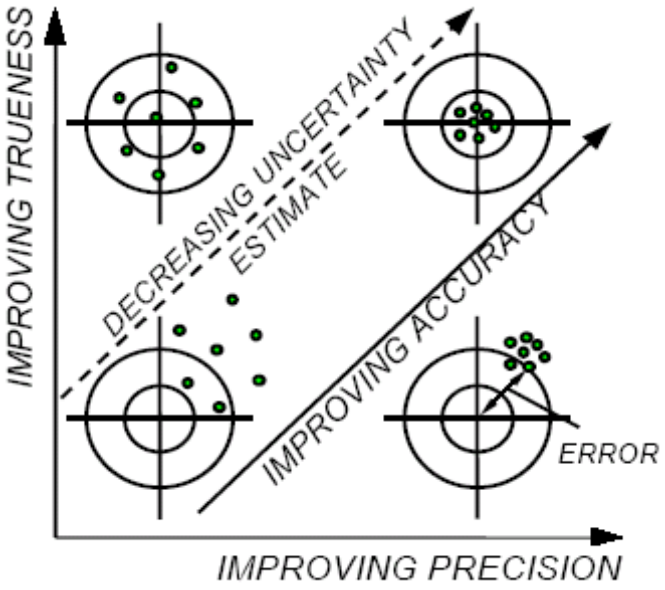


Fig 1. Relationships between trueness, accuracy, precision and uncertainty in analytical results (AMC, 2003). (Reproduced by permission of the Royal Society of Chemistry, London)

1.8. The concepts of accuracy and trueness must take account of error and precision. Uncertainty estimates (*qv*) provide a simple way to quantify such needs. However, since in a real-life situation we never know what the "true" or "correct" answer is, trueness can be assessed only in a validation-type trial against an accepted reference value. This is much more complex in microbiology than it is in physics, and chemistry.

2. Uncertainty of Measurement

2.1. The ISO/Eurachem (2000) definition of **Uncertainty of a Measurement** is

2.2. "*A parameter associated with the result of a measurement that characterises the dispersion of the values that could reasonably be attributed to the measurand*". The term "measurand" is a bureaucratic way of saying "analyte".

2.3. Translated into simple English this definition can be rewritten, as "*Uncertainty is a measure of the likely range of values that is indicated by an analytical result*."

2.4. For quantitative data (e.g. colony counts, MPNs or LOD_{50} values) a measure of uncertainty may be any appropriate statistical parameter associated with the test result. Such parameters include the standard deviation, the standard error of the mean or a confidence interval around that mean.

2.5. Measures of repeatability and reproducibility are the corner stones of estimation of analytical uncertainty. They are defined (ISO 2004) as:

2.5.1. *Repeatability* is "a measure of variability derived under specified ***repeatability conditions"***, i.e. independent test results are obtained with the same method on identical test items in the same laboratory by the same analyst using the same equipment, batch of culture media and diluents, and tested within short intervals of time.

2.5.2. *Reproducibility* is "a measure of precision derived under ***reproducibility conditions"*** i.e. test results are obtained with the same method on identical test items in different laboratories with different operators using different equipment. A valid statement of reproducibility requires specification of the conditions used.

2.5.3. **Intermediate Reproducibility** (ISO 5725-2:1994) is defined as "a measure of reproducibility derived under reproducibility conditions within a single laboratory".

2.5.4. **Standard Uncertainty** of a measurement ($u(y)$) is defined (GUM, 2000) as "the result obtained from the values of a number of other quantities, equal to the positive square root of a sum of terms, the terms being the variances or covariances of these other quantities weighted according to how the measurement result varies with changes in these quantities"

2.5.5. **Expanded Uncertainty** (U) is defined as "the quantity defining an interval about a result of a measurement expected to encompass a large fraction of the distribution of values that could reasonably be attributed to the measurand".

2.5.6. The "*Expanded Uncertainty"* values are derived by multiplying the SD's with a "*coverage factor*" to provide confidence intervals for repeatability and reproducibility around the mean value. Routinely, a coverage factor of 2 is used to give approximate 95% distribution limits (confidence interval) around the "normalised" mean value.

2.6. For qualitative data (e.g. presence or absence tests) uncertainty measures cannot be derived in the same way. However, other procedures e.g. use of the standard error associated with derived values for e.g. LOD_{50} (*qv*) and by binomial analysis of the relative proportions of positive and negative results in a comparative evaluation of methods (see 3.4 below).

3. How is uncertainty estimated?

3.1. There are two totally different approaches to the estimation of uncertainty:

3.1.1. The "bottom up" approach in which the errors associated with **all** the relevant steps undertaken during an analysis are used to derive a value for the "*combined standard uncertainty*" associated with a method (Eurachem 2000; Niemelä, 2002). Essentially this approach provides a broad indication of the possible level of uncertainty associated with method rather than a measurement; ISO TC34 SC9 considers the approach always to under-estimate the extent of variation since it cannot take into account either matrix-associated errors or the actual day-to-day variation seen in a laboratory. For

these reasons, ISO has recommended that this approach is not appropriate for microbiological analyses.

3.1.2. The "top-down" approach is based on statistical analysis of data generated in intra- or inter-laboratory collaborative studies on the use of a method to analyze a diversity of matrixes. It therefore provides *an estimate of the uncertainty of a measurement* associated with the use of a specific method.

3.1.3. Statistical aspects of the procedures, together with worked examples, for both approaches are summarised in Annexes I & II.

3.1.4. A review of measurement uncertainty in quantitative microbiological analysis is currently in press (Corry et al, 2006).

3.2. ***Quantitative Tests.*** For quantitative data (e.g. colony counts and MPN estimates), measures of "repeatability" and "reproducibility" are derived as the standard deviations of repeatability (s_r) and reproducibility (s_R). However…

3.2.1. Microbiological data do not normally conform to a "normal" distribution, and usually require mathematical transformation prior to statistical analysis. For most purposes, a $\log_{10}$ transformation is used to "normalise" the data but in cases of significant over-dispersion the use of a negative-binomial transformation may be necessary (Jarvis, 1989; Niemelä, 2002). If there is reason to believe that data conform to a Poisson distribution, then a square root transformation is required, since the variance (σ^2) is numerically equal to the mean (m) value.

3.2.2. Statistical analyses of collaborative trial data are generally done by Analysis of Variance (ANOVA) after removing any outlying values, as described by Youden & Steiner (1975) and by Horwitz (1995). However, it has been argued (e.g. AMC 1989, 2001) that it is wrong to eliminate outlier data and that application of Robust Methods of analysis is preferable.

3.2.3. One approach to robust analysis is a "robusticised" ANOVA procedure based on Huber's H15 estimators for the robust mean and standard deviation of the data (AMC, 1989, AMC 2001, ISO 5725-5:1998).

3.2.4. An alternative approach is that of the Recursive Median (REMEDIAN) procedure (ISO 2000; Wilrich, 2005).

3.2.5. Worked examples of traditional and robust analyses are shown in Annexe III.

3.2.6. A major drawback to use of these robust techniques for inter-laboratory trials is that they do not permit the derivation of Components of Variance. A novel approach to overcome this disadvantage is by the use of stepwise robust analysis for "nested" trial data, as described by Hedges & Jarvis (2006).

3.3. ***Intermediate Reproducibility of Quantitative Tests.*** Similar procedures may be used to estimate intermediate (intra-laboratory) reproducibility associated with the use of an analytical procedure in a single laboratory. Even data obtained, for instance, in laboratory quality monitoring can be used to provide an estimate of intra-laboratory

reproducibility. ISO/PTDS 19036:2005 (Part 6) describes a statistical procedure for analysis of paired data. A worked example is shown in Annex IV.

3.4. ***Qualitative Tests***. Estimation of uncertainty associated with qualitative (e.g. presence or absence) methods has not been well documented and is currently the subject of discussion within ISO.

3.4.1. Many of the potential errors that affect quantitative methods also affect qualitative methods; but there are also some additional potential errors that are inherent in the analytical procedure. For example:

3.4.1.1. In taking a sample for analysis, it is of critical importance to have knowledge of the probable distribution of organisms in the test matrix, especially when testing for organisms at the limit of detection of a method. Whilst it *may* be possible to ensure reasonable conformity with a Poisson (random) distribution of index organisms in artificial test matrixes, such distribution should not be assumed to occur in natural matrixes and requires confirmation (e.g. using an Index of Dispersion Test such as that described by Fisher et al, 1922) before using such matrixes in collaborative studies. In real life testing, erroneous decisions can result from an assumption that all microorganisms are distributed randomly at low level – there are some well-documented examples where "over dispersion" of organisms (e.g. due for instance to clumping) has resulted in a significant level of genuine false negative surveillance data.

3.4.1.2. There is an intrinsic need to ensure effective growth of the index organism to critical levels during all pre-enrichment, enrichment and differential/diagnostic culture stages – so culture medium composition, incubation times & temperatures, etc are critical to the success of the test.

3.4.1.3. It is critical to ensure that the confirmatory stages of a test protocol do actually identify the index organism.

3.4.1.4. Knowledge of the potential effect of competitive organisms is of major importance for all cultural and confirmatory stages of a test protocol.

3.4.1.5. The decision on use of either true pairs or non-paired samples is of great importance in the interpretation of potential false negative or false positive results for method validation studies.

3.4.2. The output of qualitative tests is a series of positive and negative responses. One approach to seeking to quantify such data was the derivation of the Accordance and Concordance concept (Langton et al, 2002) that sought to provide measures "equivalent to the conceptual aspects of repeatability and reproducibility". However, it is now considered that this approach is not sufficiently robust to be used in the manner proposed and adds no value to the original data.

3.4.3. Provided that a sufficient number of parallel tests has been undertaken at each of several levels of potential contamination, then it is possible to quantify the

test responses in terms of an estimated Level of Detection for (e.g.) 50% positives [LOD_{50}](for details see Hitchins, 2005).

3.4.3.1. This statistical approach essentially estimates the Most Probable Number of organisms at each test level and then analyses the relative MPN values using the Spearman-Karber approach.

3.4.3.2. Alternative approaches including Probit and Logit analyses may also be appropriate in specific circumstances.

3.4.3.3. What these methods have in common is an ability to transform purely qualitative data into a quantitative format for which error values can be derived so permitting an estimate of the uncertainty of the test result.

3.4.3.4. An extrapolation of the approach would be to determine also the LOD_0 and LOD_{90} values such that a dose-response curve can be derived. This may be of importance in differentiating between methods capable of detecting specific organisms at a similar LOD_{50} level but for which the absolute limit of non-detection (LOD_0) and a selected higher limit of detection (e.g. LOD_{90}) differ.

3.4.3.5. An alternative approach is to estimate the uncertainty associated with the proportions of test samples giving a positive response, based on the binomial distribution.

3.4.4. Examples of the way in which such approaches to analysis of qualitative data can be used are illustrated in Annex V.

4. Reporting of Uncertainty

4.1. The expression of uncertainty is of some importance in interpretation of data. Assuming a mean aerobic colony count (ACC) = 5.00 (log_{10}) cfu/g and a reproducibility standard deviation of ± 0.25 (log_{10}) cfu/g, then the expanded uncertainty is given, for instance, by:

4.1.1. Aerobic colony count on product X is 5.00 ± 0.50 (log_{10}) cfu/g; or

4.1.2. Aerobic colony count on product X is 5.00 (log_{10}) cfu/g ± 10%

4.2. It is important not to refer to analytical methods as having a precision of e.g. ± 10% based on uncertainty estimates. Uncertainty is a measure of variability i.e. a measure of the lack of precision.

5. The use of uncertainty measures in assessing compliance of a test result with a defined criterion is of some importance and has been considered by the European Commission (Anon, 2003). Jarvis et al (2004) and Jarvis & van der Voet (2005) have discussed the interpretation of data in relation to microbiological criteria for foods.

For more information, please contact Basil Jarvis at basil.jarvis@btconnect.com.

6. References cited

Analytical Methods Committee (1989) Robust Statistics – How Not to Reject Outliers. Part 1: Basic concepts, *The Analyst* **114**, 1693 – 1697. Part 2: Inter-laboratory trials, *The Analyst* **114**, 1699 – 1702.

Analytical Methods Committee (2001) Robust statistics: a method of coping with outliers. AMC Brief No.6, Royal Society of Chemistry, London.

Analytical Methods Committee (2003) Terminology - the key to understanding analytical science. Part 1: Accuracy, precision and uncertainty. AMC Brief No.13, Royal Society of Chemistry, London.

Anon (2003) The relationship between analytical results, the measurement uncertainty, recovery factors and the provisions in EU food and feed legislation. Report to the EU Standing Committee on the Food Chain and Animal Health Working Group Draft, 5 June 2003

Corry, J, Jarvis, B, Passmore, S and Hedges, A (2006) A critical review of measurement uncertainty in the enumeration of food microorganisms. *Food Microbiology* (in press)

Eurachem (2000) *Quantifying Uncertainty in Analytical Measurement*. 2nd edition, Laboratory of the Government Chemist, London.

Fisher, R A , Thornton, H G & Mackenzie, W A (1922) The accuracy of the plating method of estimating the density of bacterial populations. *Annals Applied Biology,* **9**, 325 – 359.

Hedges, A & Jarvis, B (2006) Application of robust methods to the analysis of collaborative trial data using bacterial colony counts.. *J Microbiological Methods (in press).*

Hitchins, A J (2005) Proposed Use of a 50 % Limit of Detection Value in Defining Uncertainty Limits in the Validation of Presence-Absence Microbial Detection Methods. *BPMM Report for the Statistics WG* (010705).

Horwitz, W (1995) Protocol for the design, conduct and interpretation of method performance studies. *Pure & Applied Chemistry*, **67**, 331 – 343.

ISO 5725-2:1994: Accuracy (trueness and precision) of measurement methods and results – Part 2: Basic method for the determination of repeatability and reproducibility of a standard measurement method. *International Standardisation Organisation, Geneva*

ISO 5725-5:1998: Accuracy (trueness and precision) of measurement methods and results – Part 5: Alternative methods for the determination of the precision of a standard measurement method. *International Standardisation Organisation, Geneva*

ISO FDIS 16140:2000 Microbiology *of food and* animal *feeding stuffs – Protocol for the validation of alternative methods. International Standardisation Organisation, Geneva.*

ISO 3534-1:2003 Statistics – Vocabulary and Symbols. *International Standardisation Organisation, Geneva.*

ISO 16140:2003 Microbiology of food and animal feeding stuffs — Protocol for the validation of alternative methods. *International Standardisation Organisation, Geneva.*

ISO TS 21748:2004 Guidance for the use of repeatability, reproducibility and trueness estimates in measurement uncertainty estimation *International Standardisation Organisation, Geneva*

ISO PTDS 19036:2005 Microbiology of food and animal feeding stuffs –Guide on estimation of measurement uncertainty for quantitative determinations*. International Standardisation Organisation, Geneva*

Jarvis, B (*1989*) *Statistical analysis of the microbiological analysis of foods.* Progress in Industrial Microbiology Vol 21. Elsevier, Amsterdam.

Jarvis, B (2000) Sampling for Microbiological Analysis. In The Microbiological Safety and Quality of Food. Ed B M Lund, A C Baird-Parker & G W Gould. Vol II, pp.1691 – 1733. Aspen Pub Inc, Gaithersburg, MA.

Jarvis, B, Hedges, A, Corry, J E L & Wood, R (2004) Certainty or Uncertainty? The impact of uncertainty on the interpretation of colony count data in relation to microbiological criteria*. Poster presented at the Society for Applied Microbiology conference on "Food and Dairy Microbiology", Cork, July 2004.*

Jarvis, B & van der Voet, H (2005) Guidelines on the use of uncertainty measurements in the assessment of data for compliance with quantitative microbiological criteria for foods. Draft working paper for ISO TC34 SC9 Statistics WG.

LaBarre, D D, Zelenke, D & Flowers, R (2005) Intra-laboratory and Inter-laboratory Variability. *BPMM Sampling WP document* (draft 5 – 6/9/05)

Langton, S.D, Chevennement, R, Nagelkeke, N, Lombard, B. (2002). Analysing collaborative trials for qualitative microbiological methods: accordance and concordance. *International Journal of Food Microbiology* **79**, 171-181.

Niemelä, S.I, 2002 Uncertainty of quantitative determinations derived by cultivation of microorganisms. 2nd edition, Centre for Metrology and Accreditation, Advisory Commission for Metrology, Chemistry Section, Expert Group for Microbiology, Helsinki, Finland, Publication J3/2002

Niemelä, SI (2003) Measurement uncertainty of microbiological viable counts. *Accreditation and Quality Assurance*, 8: 559-563.

NMKL (2002). Measurement of uncertainty in microbiological examination of foods. NMKL *Procedure no 8, 2nd ed., Nordic Committee on Food Analysis*

Rousseeuw, P. J., and Croux, C. (1993): Alternatives to the median absolute deviation. *J. Am. Stat. Ass*. **88**, 1273-1283

SMITH, P.A. & KOKIC, P. (1996) Winsorisation in ONS business surveys. *Working paper no. 22 at the UN Data Editing Conference 1996, Voorburg.*

Van der Voet, H (2005) Alternative models for measurement uncertainty of microbiological count data. Paper presented to ISO TC34 SC9 Statistics WG meeting, Parma, April 2004.

VIM (1993) International vocabulary of basic and general terms in metrology. (ISO 1993)

Wilrich, P-T (2005a) Robust estimates of the theoretical standard deviation to be used in inter-laboratory precision experiments. Discussion paper for the ISO TC34 SC9 Statistics Group Meeting, Paris, April 2005.

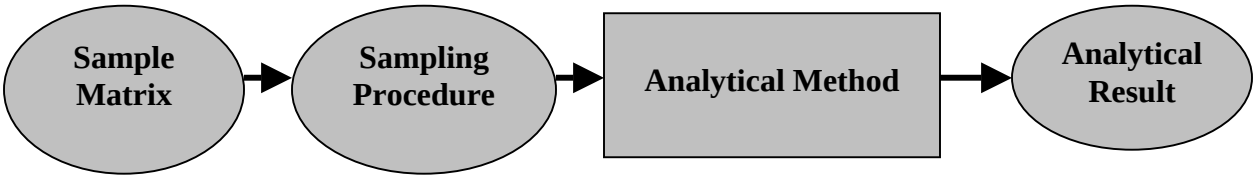
Wilrich, P-T (2005b) The determination of precision of measurement methods with qualitative results by interlaboratory experiments. Discussion paper for the ISO TC34 SC9 Statistics Group Meeting, Paris, April 2005.

Youden, W J and Steiner, E H (1975) *Statistical Manual of the AOAC*. AOAC, Washington.

Annex I

Top-Down Procedure For Estimation Of Uncertainty

1. The basis of the "top down" approach described by GUM (Eurachem 2000) is to identify and take account of all procedural stages of an analytical method. The variance associated with each individual stage is combined with the variances all the other stages and interactions that make up an analytical procedure in order to estimate a generic level of uncertainty for a method. This is illustrated diagrammatically in the schematic below.



2. Consider first the sample matrix: what are the likely errors that will affect the analytical result?

2.1. The largest potential error sources will be: the spatial distribution of the microorganisms (random, under- or over dispersion as exemplified by evidence of clumping); the condition of the microorganisms (viable and vital, sublethally damaged, non-cultivatable); the effects of competitive organisms on the recoverability of specific types; whether the organisms are located primarily on the surface of, or more generally distributed throughout, the matrix; etc.

2.2. However, the intrinsic nature of the matrix will also affect the results of an analysis.

3. How representative is an analytical sample taken from a matrix?

3.3. Should the analytical sample be totally representative of the whole matrix, or should it relate only to a specific part, e.g. the surface of a meat carcass? If the former should the matrix be homogenized prior to taking a sample; if the latter should the surface layer be excised, swabbed, rinsed or tested using a replica plating technique? What ever the method of sampling to what extent is the microflora in the analytical sample representative of both the number and types of microorganisms present in the original matrix.

3.4. If the matrix is a composite food, should the sample represent the whole or individual parts of the food matrix (e.g. in the case of a meat pie should the pastry and the meat be analysed separately)?

3.5. What size of sample should be tested? Increasing the size of an analytical sample results in a decrease in the standard error associated with the mean weight of sample taken. Similarly, increasing the weight of sample taken tends to increase the apparent colony count whilst reducing the overall variance of the mean count (Jarvis, 1989).

4. At its simplest, the analytical process consists of taking an analytical sample, suspending that sample in a defined volume of a suitable primary diluent, macerating the sample, preparing serial dilutions, plating measured volumes onto or into a culture medium, incubating the plates, counting and recording the numbers of colonies and deriving a final estimate of colony forming units (cfu) in the original matrix. At all stages throughout this process, errors will occur.

4.1. Some errors, e.g. those associated with the accuracy of weighing, the accuracy of pipette volumes, the accuracy of colony counting, etc, etc can be quantified and measures of the variance can be derived.

4.2. Some errors can be assessed, but not necessarily quantified; for instance, laboratory quality control procedures can be used to assess the extent to which a culture medium will support the growth of specific organisms. Such data may potentially provide a correction factor for the yield of organisms on a particular culture medium; whether or not the use of a correction factor should be employed in microbiological practice is a matter of debate!

4.3. However, other errors, such as those associated with individual technical performance on a day, cannot be quantified.

5. Some analytical errors associated with microbiological practices are possibly not significant when compared to other errors, but how do you know this if the errors cannot be quantified? To assess the uncertainty of an analytical microbiological procedure from the "top down" requires a full evaluation of **all** potential sources of error for each and every stage of an analytical procedure.

6. Estimation of the standard uncertainty of an analytical procedure, once a reliable schedule of quantifiable errors has been produced, is done simply by combining the errors:

$$s_R^2 = s_a^2 + s_b^2 + ... + s_x^2 + s_y^2 + s_z^2$$

where s_R^2 = reproducibility variance of the method and $s_{a....z}^2$ = variance of any stage (a….z) within the overall method.

By definition, the reproducibility standard deviation (s_R) is derived from the square root of the variance:

$$s_R = \sqrt{s_a^2 + s_b^2 + ... + s_x^2 + s_y^2 + s_z^2}$$

7. The expanded uncertainty is derived by multiplying the standard uncertainty by a coverage factor *k*, which has a value from 2 to 3. A value of 2 is normally used to give approximate 95% confidence limits; hence

$$U = k.\ s_R = 2.\ s_R$$

8. Niemelä (2002, 2003) gives a more detailed explanation of the "top down" approach to assessment of measurement uncertainty in microbiological analysis.

Annex II

"Bottom-up" Approach to Estimation of Uncertainty

1. Traditionally, the parameters used to derive uncertainty measures are estimated from the pooled results of a "valid" inter-laboratory collaborative study, or in the case of intermediate reproducibility, from an intra-laboratory study. Appropriate procedures to ensure that the study design is valid have been described *inter alia* by Youden & Steiner (1975) and by ISO (1994, 1998).

2. The data from all participating laboratories are subjected to analysis of variance (ANOVA) after first checking for:

2.1. Conformance with a "normal distribution" either by plotting the data or by application of appropriate tests for "normality".

2.2. Identification and removal of "outliers" using the methods described by Youden and Steiner (1975) or Horwitz (1995), followed if necessary by repeating the tests for conformance with "normality".

3. Quantitative microbiological data (e.g. colony counts and MPNs) do not conform to a normal distribution and require transformation to "normalise" the data before analysis.

4. Transformations are done by converting each of the raw data values (x_i) into the $\log_{10}$ value (y_i) where $y_i = \log_{10} x_i$. Strictly, it is more correct to use the natural logarithmic transformation (i.e. $y_i = \ln x_i$) (van der Voet, 2004).

5. For low level counts (typically < 100 cfu/g) that conform to the Poisson distribution (mean value (m) = variance (s^2)), the data are transformed by taking the square root of each data value (i.e. $y_i = \sqrt{x_i}$).

6. However, because of problems of over-dispersion frequently associated with microbial contamination, it may be preferable to test for (or to assume) conformance with a negative binomial distribution. Some statistical packages (e.g. Genstat) include a facility to make this transformation (using the Maximum Likelihood Method programme RNEGBINOMIAL), but such procedures are not universally available and it can be very time-consuming to calculate manually (Jarvis, 1989; NMKL, 2002, Niemelä, 2003; van der Voet, 2004).

7. Assuming a fully "nested" experimental design (e.g. duplicate testing of duplicate samples by "*A*" analysts in each of "*L*" laboratories), the residual mean variance (i.e. the variance of the replicated analyses on each sample) of the ANOVA provides an estimate of repeatability variance (s_r^2). The estimate of reproducibility variance (s_R^2) first requires computation of the contributions to variance of the samples, analysts and laboratories. This is illustrated below.

8. The repeatability standard deviation (s_r) and the reproducibility standard deviation (s_R), being the square root values of the respective variances, are the measures of standard uncertainty from which the expanded uncertainty estimates are derived.

9. ***Statistical Procedure to Derive Component Variances from an ANOVA Analysis***

Assume: trial consists of (*p*) laboratories (*p=20*) in each of which 2 analysts test 2 replicate samples and make duplicate analyses of each sample. Hence, each laboratory carries out 8 replicate analyses and the total number of analyses = 8p = 160.

Each data value (y_{pijk}) is allocated to a cell in the data table in the sequence laboratory (*p*), analyst (*i*), sample (*j*) and replicate (*k*), as shown below, and are then analysed by multivariate analysis of variance.

Laboratory (*p* = *1…20*)	Analyst (*i =1*)				Analyst (*i = 2*)			
	Sample (*j = 1*)		Sample (*j = 2*)		Sample (*j = 1*)		Sample (*j = 2*)	
	Replicate (*k = 1*)	Replicate (*k = 2*)	Replicate (*k = 1*)	Replicate (*k = 2*)	Replicate (*k = 1*)	Replicate (*k = 2*)	Replicate (*k = 1*)	Replicate (*k = 2*)
1	y_{1111}	y_{1112}	y_{1121}	y_{1122}	y_{1211}	y_{1212}	y_{1221}	y_{1222}
2	y_{2111}	y_{2112}	y_{2121}	y_{2122}	y_{2211}	y_{2212}	y_{2221}	y_{2222}
3	y_{3111}	y_{3112}	y_{3121}	y_{3122}	y_{3211}	y_{3212}	y_{3221}	y_{3222}
4	y_{4111}	y_{4112}	y_{4121}	…	…	…	…	…
…	…	…	…	…	…	…	…	…
…	…	…	…	…	…	…	…	…
20	y_{20111}	y_{20112}	y_{20121}	y_{20122}	y_{20211}	y_{20212}	y_{20221}	y_{20222}

ANOVA table for a four-factor fully-nested experiment

Source of Variation	Sum of Squares	Degrees of freedom	Mean Square	Expected Mean Square Components*
Laboratories	SS_{lab}	p-1 = 19	$SS_{lab}/19 = MS_{lab}$	$\sigma_r^2 + 2\sigma_{sam}^2 + 4\sigma_{ana}^2 + 8\sigma_{lab}^2$
Analysts	SS_{ana}	p = 20	$SS_{ana}/20 = MS_{ana}$	$\sigma_{\boldsymbol{r}}^2 + 2\sigma_{\boldsymbol{sam}}^2 + 4\sigma_{\boldsymbol{ana}}^2$
Samples	SS_{sam}	2p = 40	$SS_{sam}/40 = MS_{sam}$	$\sigma_r^2 + 2\sigma_{sam}^2$
Residual	SS_{res}	4p = 80	$SS_{res}/80 = MS_{res}$	σ_r^2
Total	Total SS	8p –1=159		

* The components are shown as population variances since this is an expectation table.

The residual mean square ($MS_{res} = s_r^2$) provides the repeatability variance between duplicate analyses done on the same replicate sample.

The variance due to l samples (s^2_{sam})is given by $[MS_{sam} - s^2_r]/2$

The variance due to analysts (s^2_{ana}) is given by $[MS_{ana} - 2s^2_{sam} - s^2_r]/4$

The variance due to is laboratories (s^2_{lab}) given by $[MS_{lab} - 2s^2_{sam} - 4s^2_{ana} - s^2_r]/8$

The Reproducibility Variance (s^2_R) is given by $[s^2_{sam} + s^2_{ana} + s^1_{lab} + s^2_r]$

The Reproducibility Standard Deviation is given by $\sqrt{s^2_{sam} + s^2_{ana} + s^1_{lab} + s^2_r}$

The Repeatability Standard Deviation is given by $\sqrt{s^2_r}$.

WORKED EXAMPLE (10 Labs x 2 Analysts x 2 Samples x 2 Replicate analyses)

Log transformed colony counts (Log_{10} cfu/g)

Laboratory	Analyst (*i = 1*)				Analyst (*i = 2*)			
	Sample (*j=1*)		Sample (*j=2*)		Sample (*j=1*)		Sample (*j=2*)	
	Replicate (*k = 1*)	Replicate (*k = 2*)	Replicate (*k = 1*)	Replicate (*k = 2*)	Replicate (*k = 1*)	Replicate (*k = 2*)	Replicate (*k = 1*)	Replicate (*k = 2*)
1	5.56	5.73	5.76	5.59	6.08	5.96	6.07	5.99
2	6.02	5.88	5.87	5.80	5.54	5.63	5.92	5.79
3	6.26	6.30	6.46	6.54	6.42	6.49	6.11	6.42
4	5.07	5.11	4.90	4.61	4.63	4.81	4.42	4.56
5	5.39	5.25	5.28	5.52	5.34	5.46	5.47	5.49
6	5.98	5.88	6.02	5.64	5.96	6.06	5.70	5.57
7	5.43	5.18	5.16	5.08	6.15	5.76	5.44	5.43
8	5.94	5.73	5.28	5.47	5.99	6.01	5.92	6.13
9	5.45	5.35	5.49	5.42	5.68	5.57	5.74	5.69
10	5.51	5.74	6.18	6.13	5.83	5.91	5.76	5.60

Tests for normality (e.g. Shapiro-Wilk, W = 0.9830, p= 0.0885) did not disprove the hypothesis that the $\log_{10}$ transformed data conform reasonably (although not perfectly) to a normal distribution. However, application of the Cochran Test (Horwitz, 1995) identified Laboratory 7 as an outlier; subsequently evaluation using the Grubbs test did not eliminate other laboratories although laboratories 3 & 4 appeared to be possible outliers.

ANOVA table for the four-factor fully nested experiment (All data included)

Source of Variation	Sum of Squares	Degrees of freedom	Mean Square (rounded to 4 places)	Mean Square Components
Laboratories	12.636	9	1.4040	$s_r^2 + 2s_{sam}^2 + 4s_{ana}^2 + 8s_{lab}^2$
Analysts	1.4906	10	0.1491	$s_r^2 + 2s_{sam}^2 + 4s_{ana}^2$
Samples	1.346	20	0.0673	$s_r^2 + 2s_{sam}^2$
Residual	0.5554	40	0.0139	s_r^2
Total	16.0272	79		

The residual mean square ($MS_{res} = s_r^2 = 0.0139$) provides the repeatability variance between duplicate analyses done on the same replicate sample.

Component Variances

Sample variance (s_{sam}^2) = [MS_{sam}- s_r^2]/2 = [0.0673 – 0.01389]/2 = 0.0267

Analyst variance (s_{ana}^2) = $[MS_{ana} - 2s_{sam}^2 - s_r^2]/4$ = [0.1491 – 0.0673]/4 = 0.02045

Laboratory variance (s_{lab}^2)= $[MS_{lab} - 2s_{sam}^2 - 4s_{ana}^2 - s_r^2]/8$ = [1.4040 – 0.1491]/8 = 0.1548

Hence, Reproducibility Variance (s_R^2) = $[s_{sam}^2 + s_{ana}^2 + s_{lab}^1 + s_r^2]$ =

[0.0139 +0.0267 +0.02045 +0.15686] = 0.2179

Reproducibility Standard Deviation = $s_R = \sqrt{s_{sam}^2 + s_{ana}^2 + s_{lab}^1 + s_r^2}$ = √0.2179 = ±0.4668

Repeatability Standard Deviation = $s_r = \sqrt{s_r^2}$ = √0.01389 = ±0.1178

The mean colony count = 5.6682 ≈ 5.67 ($\log_{10}$) cfu/g

Hence, Relative Standard Deviation of Reproducibility (RSD_R)

= 100 x 0.4668/5.6682 = 8.24%

and, Relative Standard Deviation of Repeatability (RSD_r)

= 100 x 0.1178/5.6682 = 2.08%

From these values the 95% expanded uncertainty of reproducibility is given by:

$U = 2\ s_R$ = 2 x 0.4668 = ±0.9336. ≈ ±0.93 ($\log_{10}$) cfu/g

The upper and lower limits of the 95% Confidence Interval on the mean colony count are:

$U_L = 5.67 + 0.93 = 6.60$ ($\log_{10}$) cfu/g

$L_L = 5.67 - 0.93 = 4.74$ ($\log_{10}$) cfu/g

Repeat analyses for 9 laboratories(after elimination of data for laboratory 7)

Source of Variation	Sum of Squares	Degrees of freedom	Mean Square (rounded to 4 places)	Mean Square Components
Laboratories	12.227	8	1.5284	$s_r^2 + 2s_{sam}^2 + 4s_{ana}^2 + 8s_{lab}^2$
Analysts	1.0249	9	0.1139	$s_r^2 + 2s_{sam}^2 + 4s_{ana}^2$
Samples	1.0405	18	0.0578	$s_r^2 + 2s_{sam}^2$
Residual	0.4449	36	0.0124	s_r^2
Total	16.0272	71		

The component variances were derived as:

Repeatability variance (s_r^2) = 0.0124 Sample variance (s_{sam}^2) = 0.0227

Analyst variance (s_{ana}^2) = 0.0140 Laboratory variance (s_{lab}^2) = 0.1168

Hence, Reproducibility Variance (s_R^2) = 0.2279

Reproducibility Standard Deviation = $s_R = \sqrt{0.2279} = \pm 0.4753$

Repeatability Standard Deviation = $s_r = \sqrt{0.0124} = \pm 0.1112$

The mean colony count = 5.6921 ≈ 5.69 ($\log_{10}$) cfu/g

Hence, Relative Standard Deviation of Reproducibility (RSD_R) = 8.35%

and, Relative Standard Deviation of Repeatability (RSD_r) = 1.95%

From these values the 95% expanded uncertainty of reproducibility is given by:

$U = 2\ s_R = 2 \times 0.4753 = \pm 0.9506. \approx \pm 0.95$ ($\log_{10}$) cfu/g

The upper and lower limits of the 95% Confidence Interval on the mean colony count are:

$U_L = 5.69 + 0.95 = 6.64$($\log_{10}$) cfu/g

$L_L = 5.67 - 0.95 = 4.72$ ($\log_{10}$) cfu/g

Comparison of ANOVAs with and without removal of outlier laboratory

The table below shows that removal of one set of data (from the outlier laboratory) marginally increased the mean colony count and reduced the component variances for repeatability, samples, analysts and laboratories. However the overall effect, in this specific example, was marginal in relation to the derived values for repeatability and reproducibility; and hence there was little effect on the level of expanded uncertainty.

Parameter	10 Laboratories	9 Laboratories
Mean Colony Count (log_{10} cfu/g)	5.6682	5.6921
Repeatability Variance	0.0139	0.0124
Sample Variance	0.0267	0.0227
Analyst Variance	0.0205	0.0140
Laboratory Variance	0.1548	0.1168
SD repeatability (SD_r)	±0.1178	±0.1112
Relative SD_r	2.08%	1.95%
SD reproducibility (SD_R)	±0.4668	±0.4753
Relative SD_R	8.24%	8.35%
Expanded Uncertainty (U)	±0.93	±0.95
Upper Limit of 95% CI (log_{10} cfu/g)	6.60	6.64
Lower Limit of 95% CI (log_{10} cfu/g)	4.74	4.72

Annex III

Estimation of Intermediate Reproducibility based on Routine Monitoring Data

1. Intra-laboratory uncertainty estimates can be made either by carrying out a full internal collaborative trial, with different analysts testing the same samples over a number of days or, for instance, using different batches or even different brands of commercial culture media. In such a case the statistical procedure of choice is that described in Annex II.

2. However, if a laboratory undertakes routine quality monitoring tests, it is possible to estimate reproducibility from these test data. One approach is to use a 1-way ANOVA and to take the mean residual square as the estimate of reproducibility. A preferred, and simpler procedure, is described fully in ISO19036: 2005; this determines the variance for each set of transformed replicate data values.

3. The reproducibility standard deviation is derived from the square root of the sum of the duplicate variances divided by the number of data sets. The equation is:

$$S_R = \sqrt{\sum_{i=1}^{n} \frac{(y_{i1} - y_{i2})^2 / 2}{n}}$$

where y_{i1} and y_{i2} are the log transformed values of the original duplicate counts (x_1 and x_2) and n is the number of pairs of counts.

4. A worked example (based on $\log_{10}$ transformation) is presented below.

5. Confusion sometimes arises between repeatability and intermediate reproducibility. It must always be remembered that repeatability requires all stages of the replicated tests to be done **only** by a single analyst, carrying out repeat determinations on a single sample in a single laboratory, using identical culture media, diluents, etc within a short time period e.g. a few hours. If more than one analyst undertakes the analyses and/or tests are done on different samples and/or on different days then the calculation derives a measure of intermediate reproducibility. The procedure can be used to determined average repeatability estimates for individual analysts provided all the repeatability criteria are met..

6. Internal laboratory quality management is aided by the use of statistical process control (SPC). The estimates of intermediate reproducibility provide a source of data that is amenable to SPC.

Worked Example (modified from ISO 19036:2005)

The data below were derived from enumeration of aerobic mesophilic flora in mixed poultry meat samples. The duplicate data values (x_{iA} and x_{iB}) are log transformed to give y_{iA} and y_{iB}, respectively. The mean $\log_{10}$ counts ($\bar{y}$) are derived from $(y_{iA} + y_{iB})/2$; the variances (S_{Ri}^2) are derived from $(y_{iA} - y_{iB})/2$; and the RSD values from $100*\sqrt{S_{Ri}^2}/\bar{y}$.

Test(*i*)	*Colony Count A (cfu/g)*	*Colony Count B (cfu/g)*	*Log count A*	*Log count B*	*Mean log Count*	*Absolute Difference in log count*	Variance	Relative Standard Deviation (%)
	x_{iA}	x_{iB}	$y_{iA}=\log_{10}(x_{iA})$	$y_{iB}=\log_{10}(x_{iB})$	$\bar{y}$	$y_{iA} - y_{iB}$	S_{Ri}^2	RSD_{Ri}
i=1	6.70E+04	8.70E+04	4.83	4.94	4.88	0.11	0.00643	1.64%
i=2	7.10E+06	6.20E+06	6.85	6.79	6.82	0.06	0.00173	0.61%
i=3	3.50E+05	4.40E+05	5.54	5.64	5.59	0.10	0.00494	1.26%
i=4	1.00E+07	4.30E+06	7.00	6.63	6.82	0.37	0.06717	3.80%
i=5	1.90E+07	1.70E+07	7.28	7.23	7.25	0.05	0.00117	0.47%
i=6	2.30E+05	1.50E+05	5.36	5.18	5.27	0.19	0.01723	2.49%
i=7	5.30E+08	4.10E+08	8.72	8.61	8.67	0.11	0.00622	0.91%
i=8	1.00E+04	1.20E+04	4.00	4.08	4.04	0.08	0.00313	1.39%
i=9	3.00E+04	1.30E+04	4.48	4.11	4.30	0.36	0.06595	5.98%
i=10	1.10E+08	2.20E+08	8.04	8.34	8.19	0.30	0.04531	2.60%
Σ							**0.2193**	
Average					**6.18**		**0.0219**	

Using the $\log_{10}$-transformed data (y_{ij}), the reproducibility standard deviation is derived from:

$$S_R = \sqrt{\frac{\sum_{i=1}^{n}(y_{i1} - y_{i2})^2 / 2}{n}} = \sqrt{\frac{0.00643 + 0.00173 + ... + 0.04531}{10}} = \sqrt{0,02193} \approx \pm 0,15 \ \log_{10} \text{ cfu/g}$$

Average % Relative Standard Deviation (RSD_{av})= $100*(S_R/\bar{y})$=100*(0.15/6.18)=2.39%

Individual tests (i = 1….10) gave RSD values ranging from 0.47% to 5.98%, with an overall value of 2.39%.

Note: it is incorrect to take the average of the individual RSD values.

Once sufficient data are available, a moving RSD_{av} can be determined and used in a statistical process control system.

Annex IV

Application of Robust Methods of Statistical Analysis

1. Because of the problems with the occurrence of outlier data, several alternative approaches to the Analysis of Variance have been developed, based on Robust Methods of Statistical Analysis.

2. Rather than relying on identification and removal of outlying data (which values could actually be valid results, albeit considerably different from most of the data) and then estimating the variance around the mean, alternative robust procedures rely on estimation of the variation around the median value.

3. A mean value will be affected significantly by one or more high (outlier) values within a data set, whereas the median value is not affected. Consider the following examples:

A. 1, 4, 3, 6, 3, 5, 6, 3, 4, 5 $n = 10$, $\sum = 40$, Mean = 4.0 Median = 4.0

B. 1, 4, 3, 6, 3, 5, **26**, 3, 4, 5 $n = 10$, $\sum = 60$, Mean = 6.0, Median = 4.0

C. 1, 4, 3, 6, 3, 5, **26**, 3, 4, **15** $n = 10$, $\sum = 70$, Mean = 7.0, Median = 4.0

D. 1, 4, 3, 6, 3, 5, **126**, 3, 4, **15** $n = 10$, $\sum = 170$, Mean = 17.0,Median = 4.0

E. 1, 4, 3, 6, 3, 5, 3, 4, $n = 8$, $\sum = 29$, Mean = 3.6, Median = 3.5

4. The presence of one or more high values (Examples B, C, D) has a significant effect on the mean value but no effect on the median value. Removal of the high outliers (E) reduces both the mean and the median values.

5. A similar effect would be seen with low value outliers. Of course, occurrence of both high and low outliers could balance out the effect on the mean.

6. There are two primary alternative techniques of robust analysis currently in use:

6.6.1. The Analytical Methods Committee of the Royal Society Chemistry (AMC 1989, 2001) describes one approach. The procedure calculates the median absolute difference (MAD) between the results and their median value and then applies Hüber's H15 method of winsorisation. Winsorisation is a technique for reducing the effect of outlying observations on data sets (for detail see Smith & Kokic, 1996). The procedure can be used with data that conform approximately to a normal distribution but with heavy tails and/or outliers. An example is shown below. The procedure is not suitable for multimodal or heavily skewed data sets. The AMC website[1] provides downloadable software for use either in Minitab or Excel (97 or later version).

6.1.1. An alternative approach, known as the Recursive Median is based on extrapolation of the work of Rousseeuw & Croux (1993). One version of this approach

[1] www.rsc.org/lap/rsccom/amc/amc_software.htm#robustmean

(described fully in ISO 16140:2003) uses Rouseeuw's recursive median S_n. However, Wilrich (2005a) recommends a modified approach to this procedure also based on Rousseeuw's S_n computation.

Worked Example - Analysis of data set containing outliers

Duplicate Series of Colony Counts (as Log_{10} cfu/g) done by 1 Analyst in each of 10 Laboratories		
Laboratory	A	B
1	4.83	4.94
2	**4.05**	**3.99**
3	**6.84**	**6.92**
4	4.90	4.93
5	5.28	5.23
6	4.86	4.72
7	5.62	5.51
8	4.50	4.68
9	5.48	5.11
10	5.04	5.34

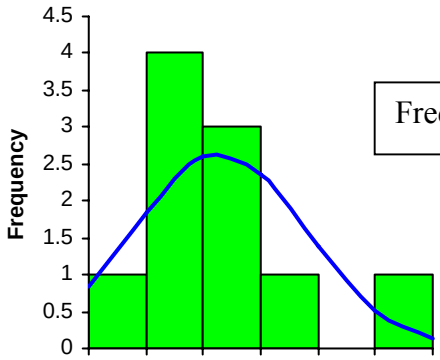
Laboratory 2 data look to be slightly low and laboratory 3 data to be high when compared with the other data.

Graphical and Descriptive Analysis of the Data

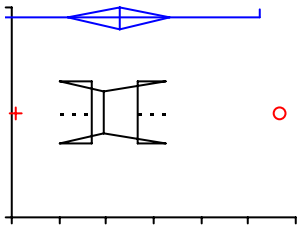
n	10
Mean	5.140
95% CI	4.602 to 5.678
Variance	0.5662
SD	0.7524
SE	0.2379
CV	15%
Median	4.970
97.9% CI	4.500 to 5.620
Range	2.79
IQR	0.49

Percentile	
2.5th	-
25th	4.838
50th	4.970
75th	5.330
97.5th	-

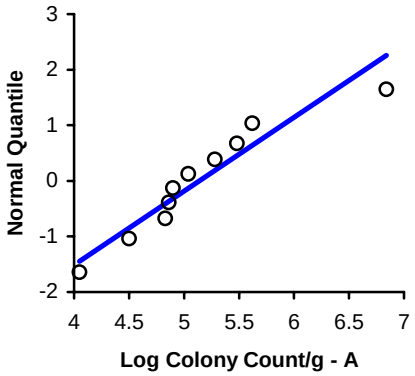
	Coefficient	**p**
Shapiro-Wilk	0.9204	0.3605
Skewness	1.1222	0.1009
Kurtosis	2.4815	-



Frequency Plot



Box Plot showing outliers



Although there is evidence of kurtosis and positive skewness, the log-transformed data conform fairly well to a "normal" distribution. The Box plot shows the presence of a potential low-level outlier (+) and a significant high-level outlier (○).

One-way Analysis of Variance (ANOVA) without removal of outliers

Source of Variation	*SS*	*df*	*MS*	*F*	*P-value*	*F crit*
Between Laboratories	10.086	9	1.1207	70.816	7E-08	3.0204
Within Laboratories	0.15825	10	0.0158			
Total	10.2443	19				

Repeatability SD = $\sqrt{0.0158} = 0.1258$
Reproducibility SD = $\sqrt{(1.1207 + 0.0158)} = \sqrt{1.1365} = 1.0661$

One-way Analysis of Variance (ANOVA) after removal of high outlier (lab 3)

Source of Variation	*SS*	*df*	*MS*	*F*	*P-value*	*F crit*
Between Laboratories	3.3464	8	0.4183	24.281	3E-05	3.2296
Within Laboratories	0.15505	9	0.0172			
Total	3.50145	17				

Repeatability SD = √0.0172 = 0.1311
Reproducibility SD = √ (0.4183+0.0172) = √0.4355 = 0.6599

One-way Analysis of Variance (ANOVA) after removal of both low and high outliers (labs 2 & 3)

Source of Variation	*SS*	*df*	*MS*	*F*	*P-value*	*F crit*
Between Laboratories	1.42124	7	0.203	10.599	0.0017	3.5005
Within Laboratories	0.15325	8	0.0192			
Total	1.57449	15				

Repeatability SD = √0.0192 = 0.1311
Reproducibility SD = √ (0.203+0.0192) = √0.2222 = 0.4714

Analysis of Variance using the AMC Method

Software for this analysis, compatible with Microsoft Excel, can be downloaded from Royal Society of Chemistry statistical software. A version for use in Minitab is also available.

ROBUST ESTIMATES

Parameter	Value
Grand Mean	5.060622
Within-laboratory/repeatability SD	0.116772
Between-laboratory SD	0.476556
Reproducibility SD	0.490654

c=1.5: Convcrit=0.0001
Repeatability SD = 0.1168
Reproducibility SD = 0.4907

Comparison of data analyses by ANOVA, without and with removal of the high (*) outlier and both the high and low outliers (**), by Robust ANOVA (AMC 1989, 2001) and by Recursive Median (ISO 16140:2003)

Parameter*	ANOVA	ANOVA*	ANOVA**	ROBUST	RECMED
Mean	5.14	4.95	5.06	5.06	
Median					5.05
SD_r	0.126	0.131	0.138	0.117	0.115
RSD_r	2.45%	2.65%	2.73%	2.31%	2.28%
SD_R	1.066	0.660	0.471	0.491	0.5590
RSD_R	20.45%	13.33%	9.42%	9.70%	11.07%

* SD_r = Standard Deviation of repeatability; SD_R = Standard Deviation of reproducibility
RSD_r = % relative standard deviation of repeatability
RSD_R = % relative standard deviation of reproducibility

The effect of the outlier values on the Standard Deviation of reproducibility is clear from the above data. Removal of the high outlier (*) reduces both the mean and the SD_R; removal of both the high and low outliers (**) reduces both the mean value and the SD_R to a level similar to than that seen in the Robust ANOVA. The Recursive Median technique (working data not shown) produces a similar value for SD_r but a somewhat higher SD_R value than does the Robust Method.

ANNEX V

Uncertainty Associated with Qualitative Methods

1. By definition, a non-quantitative method merely provides an empirical answer to a question regarding the presence or absence of a specific index organism or a group of related organisms in a given quantity of a representative sample.

2. Provided that multiple samples are analysed, and on the assumption that the test method is "perfect", then the number of tests giving a positive response provides an indication of the incidence of defective samples within a "lot".

2.1. For instance, if a test on 10 parallel samples found 4 positive and 6 negative samples then the perceived incidence of defectives would be 40% (*sic* of the samples analysed).

2.2. However, if no positive samples were found the apparent incidence of defectives in the "lot" would be zero. However, it is not possible to say that the "lot" is not contaminated because the true incidence of defective samples will be greater than zero.

3. Sampling theory for occurrence of defectives is based on the binomial distribution, in which the probability of an event occurring (p) or of not occurring (q) can be derived and an error estimate can be made based on a realistic number of samples analysed. Unfortunately, in laboratory practice it is not usually possible to analyse a realistic number of samples for the presence of specific microorganisms.

3.1. Table 1 below shows the statistical probability of occurrence of 0, 1, or 2 defective units in 10 sample units from "lots" containing from 0.1 to 30% true defectives. For a lot having only 0.1% defective units, the probability of detecting one or more defective (*sic* positive) samples is only 1 in 100 whilst for a lot having 5% true defectives there is still only a 40% probability of obtaining a positive result; even with 20% true defectives there is still a 20% chance of ***not*** finding defective units when testing 10 sample units.

3.2. Table 2 shows the probability of detecting 0, 1 or 2 defective units with increasing numbers of sample units tested when the true incidence of defectives is 10%. The probability of finding no defective samples is 59% if only 5 samples are tested, 35% with 10 samples and 12% with 20 samples.

3.3. These examples illustrate a basic characteristic of undertaking qualitative tests for specific organisms: unless the likelihood of contamination of the matrix is high, *and* the number of sample units tested is considerable *and* the analytical test itself is perfect, then the probability of detecting positive samples in food matrix is very low.

Table 1. Binomial Probability of detecting 0, 1 or 2 defective units in 10 sample units tested with increasing incidence of true defectives (mod from Jarvis, 2000)

True Incidence (%)of Defective Units in a lot	Probability (p) of detecting defective units		
	0	**1**	**2**
0.1%	0.99	0.01	<0.001
1%	0.90	0.09	<0.01
5%	0.60	0.32	0.08
10%	0.35	0.39	0.19
20%	0.20	0.35	0.28
30%	0.03	0.12	0.23

Table 2. Binomial probability of detecting defective units with increasing sample units from a lot having 10% true defectives (mod from Jarvis, 2000)

Number of Sample units (*n*) tested	Probability of detecting the following number of defective units		
	0	**1**	**2**
5	0.59	0.33	0.07
10	0.35	0.39	0.19
20	0.12	0.27	0.29
50	<0.01	0.03	0.08

3.4. ***Maximum Incidence and Level of Contamination***. Even when all test results are negative, use of the binomial distribution concept permits the derivation of a probable maximum contamination limit for a test lot.

3.4.1. Assuming that results on all (*n*) sample units are negative, then for a given probability (*p*) the maximum incidence (*d*) of defective units is given by:

$$d = 100(1 - \sqrt[n]{(1-p)}$$

Hence, if $n = 10$ and $p = 0.95$, then

$$d = 100(1 - \sqrt[10]{(1-0.95)} = 100(1 - \sqrt[n]{0.05}) = 100(1-0.741) = 25.88\% .$$

3.4.2. Knowing the maximum incidence of defective sample units and the size of the sample units we can derive a Maximum Contamination level (C) from:

$$C = (d/100)(1/W) \text{ organisms per g,}$$

where W is the weight of the sample unit tested. For the example given above and assuming that each of the 10 samples weighed 25g, then the maximum contamination level would be given by

$$C = (25.88/100)(1/25) = 0.0104 \text{ organisms/g} \equiv 10.4 \text{ organisms/Kg}$$

3.4.3. In other words, the failure to detect a positive in 10 parallel tests merely indicates, at a 95% probability, that the index organism would be present in not more than 26% of similar samples throughout the lot; and that the maximum contamination level would be 11 organisms/Kg of product.

3.4.4. It might be thought that such a level of product security is insufficient, in which case it would be necessary to analyse a greater number of sample units and ideally to test larger quantities of sample. It is essential also to recognise that this presupposes that the test method is "perfect".

3.5. ***Multiple Test Most Probable Number Estimates***. If some test results are positive, then we can derive an estimate of population density (the basis for derivation of a Most Probable Number) for multiple tests even at a single dilution level.

3.5.1. The following equation provides the derivation of the MPN:

$M = -\frac{1}{V}.ln\left(\frac{s}{n}\right)$, where M = Most probable number, V = quantity of sample, s = number of sterile tests out of n tests inoculated.

3.5.2. Assume 10 tests are set up on replicate 25g samples of product, 3 tests are positive and 7 are negative. Then the MPN of contaminating organisms is:

$$M = -\frac{1000}{25}.ln\left(\frac{7}{10}\right) = -40. - 0.3567 = 14.27 \text{ organisms/Kg} \approx 14 \text{ organisms/Kg}$$

3.5.3. Unfortunately, it is not possible to derive an estimate of the error of the MPN when tests are done at a single dilution level.

3.6. ***Level of Detection Estimates.*** The equation used in 3.5 is also the basis for deriving MPN values for use in the Spearman-Karber procedure to estimate the LOD_{50} for a test. This is the level of organisms that will give 50% positive results when tested by an appropriate protocol. Details of the procedure together with worked examples are given in the report by Hitchins (2005). This method of quantification has the benefit that it is possible to derive a value for the standard error of the mean (*sic* LOD) estimate. The procedure can be used to compare performance of two or more

methods where both have been evaluated under identical conditions in two or more laboratories.

3.7. ***Estimation of repeatability and reproducibility for qualitative tests.*** In a paper produced for ISO SC9 TC34, Wilrich (2005b) proposed the estimation of repeatability and reproducibility estimates for qualitative test procedures based on the binomial probability of detection of positive and negative results in different laboratories operating either at equal or at dissimilar sensitivity levels. A set of simulation studies is presented, together with analyses of a set of practical interlaboratory assessments, which support the proposal but the method has yet to be evaluated in detail.

4. Estimation of Error based on test performance. One of the traditional problems associated with presence or absence tests relates to the likelihood that a method may give either a false negative (Type I error) or a false positive (Type II error) result. A false negative result fails to detect the occurrence of a known index organism in a sample. A false positive result indicates the presence of a specific index organism even though it is not present in the test matrix. Such errors create specific problems for interpretation of test results.

4.1. In a real life situation, where tests are done on natural matrixes, it is impossible to estimate the likelihood of detecting such false results. It is essential therefore to ensure during development, evaluation and use of any method that the likelihood of such errors occurring is at an absolute minimum. An efficient laboratory proficiency scheme provides a way to monitor the efficiency of a test procedure in any individual laboratory.

4.2. But to be sure that false results do not occur requires the use of reference materials that can be relied upon to contain the index organism at a given level. For high-level contamination that is not a major problem; the issue arises primarily where the level of detection is intended to be close to the minimum level of detection. For instance, to detect 1 cfu of a specific organism in (say) 25g of sample implies that the organism is evenly distributed throughout a lot of test material such that each 25g sample unit is likely to contain the organism. Only if it were possible to add a single test organism to each individual 25g sample could the probability that each sample would contain that organism be achieved and even then there is the real possibility that the organism would not survive the preparation and storage process.

4.3. If larger quantities of test organism are added to a large batch of test matrix, which is then thoroughly mixed, the distribution of organisms throughout the lot would at best be random but could possibly be over dispersed due to the presence of clumps of organisms.

4.4. Table 3 shows the probability of occurrence of 0 or ≥1 organisms in a 25g sample for different levels of inoculation. If the target inoculum level is only 1 organism per 25g, there is a 37% chance that less than 1 organism will not be present in the sample; to have a 99% probability that 1 or more organisms occur in a perfectly distributed sample matrix requires inoculation at a level of at least 5 organisms/25g. Even then

one has to assume that the original inoculum contains the test organism at the relevant level – it must not be forgotten that the organisms in a well-mixed inoculum will themselves be distributed in accordance with Poisson. It is therefore perhaps not surprising that at low inoculum levels, negative results may be found frequently. It is for such reasons that we recommend the LOD_{50} approach of Hitchins (2005) for comparison of two or more methods of analysis.

5. The effect of competitor organisms and other factors on the recovery of organisms to critical levels.

5.1. A further potential cause of a false negative result is that during the multistage test protocol, the index organisms must be able to grow to a critical level to ensure effective transfer between different stages of the test. The ability of an organism to grow is dependent not only on the physiological condition of the index organism in the sample matrix, but also on the micro-environmental conditions within the test system, the presence or otherwise of competitive organisms that may affect the growth of the index organism and the time/temperature factors used in the protocol.

5.2. In their (BPMM) paper on Inter-laboratory Variability, LaBarre, Zelenka and Flowers (2005) have reviewed in detail the effects of competitive growth, problems associated with test media and other practical considerations in relation to the critical level of organisms for use in confirmatory tests following enrichment procedures. The paper describes also the statistical aspects of optimisation of test protocols.

Table 3. Probability of occurrence of 0 or at least 1 organism in a 25g sample assuming perfect random distribution of test organisms in a matrix.

Target inoculation level (cfu/25g)	Probability (p_x) of Occurrence in 25g sample units	
	<1 organism	1 or more organisms
1	0.3679	0.6321
2	0.1353	0.8647
3	0.0498	0.9502
4	0.0183	0.9817
5	0.0067	0.9933
10	<0.0001	>0.9999